

# A NONLINEAR VARIATIONAL APPROACH TO MOTION-CORRECTED RECONSTRUCTION OF DENSITY IMAGES

MARTIN BURGER<sup>†§</sup>, JAN MODERSITZKI<sup>‡¶</sup>, AND SEBASTIAN SUHR<sup>†‡</sup>

**ABSTRACT.** The aim of this paper is to establish a nonlinear variational approach to the reconstruction of moving density images from indirect dynamic measurements. Our approach is to model the dynamics as a hyperelastic deformation of an initial density including preservation of mass. Consequently we derive a variational regularization model for the reconstruction, which - besides the usual data fidelity and total variation regularization of the images - also includes a motion constraint and a hyperelastic regularization energy.

Under suitable assumptions we prove the existence of a minimizer, which relies on the concept of weak diffeomorphisms for the motion. Moreover, we study natural parameter asymptotics and regularizing properties of the variational model. Finally, we develop a computational solution method based on alternating minimization and splitting techniques, with a particular focus on dynamic PET. The potential improvements of our approach compared to conventional reconstruction techniques are investigated in appropriately designed examples.

## 1. INTRODUCTION

With current advances of imaging devices, in particular in biomedical imaging, dynamic studies including motion attract more and more attention. Prominent examples are live microscopy, MRI, and emission tomography (PET and SPECT). For the latter cardiac applications are particularly interesting, which are subject to cardiac and respiratory motion. In most of such applications a density image (e.g. the density of a radiochemical or fluorescent tracer), which is subject to motion between consecutive time steps, is to be reconstructed from indirect measurements. This problem shall be tackled with a novel variational approach in this paper.

---

<sup>†</sup> Institute for Computational and Applied Mathematics, University of Münster, Einsteinstr. 62, 48149 Münster, Germany.

<sup>§</sup> Cells in Motion Cluster of Excellence, University of Münster.

<sup>‡</sup> Institute of Mathematics and Image Computing, University of Lübeck, Maria-Goeppert-Strae 3, 23562 Lübeck, Germany.

<sup>¶</sup> Fraunhofer MEVIS Project Group Image Registration, Maria-Goeppert-Strae 3, 23562 Lübeck, Germany

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through grants BU 2327/8-1 and MO-1053/2-1, as well as by ERC via Grant EU FP 7 - ERC Consolidator Grant 615216 LifeInverse. MB acknowledges further support by the German Science Foundation DFG via EXC 1003 Cells in Motion Cluster of Excellence, Münster, Germany.

The need for advanced motion-corrected reconstruction methods is caused by the fact that separate reconstructions in small time intervals - in which the motion effect is negligible - are of inferior quality. The latter problem is either caused by severe undersampling (e.g. in MRI due to time restrictions in acquiring slice data) or the bad signal-to-noise ratio (e.g. in optical imaging and emission tomography techniques being based on counting photons) in small time steps. Several recent approaches exploiting sparsity properties of the images to be reconstructed achieve strong improvements in these two situations (cf. MR compressed sensing [39], Bregman-EM-TV for PET [44]). However, those approaches are still limited in certain practical situations, and since separate reconstruction in small time scales does not exploit natural temporal correlation, spatio-temporal image reconstruction methods are an obvious next step.

Several approaches tried to obtain improved reconstructions by appropriately averaging reconstructions at different time steps. To do so, the motion between those time steps needs to be estimated, which can be performed either by flow-type (optical flow) techniques [15] or by image registration [32]. Subsequently the motion in the image can be corrected and improved images can be obtained by averaging, cf. e.g. [32] for an implementation in a clinical PET setup. Further improvements are often achieved by iterating a reconstruction step (in separate time steps) and the motion estimation, where some motion-corrected density (initial or average) serves as a prior for the next reconstruction step. A remaining disadvantage are artefacts that can be created this way as well as unclear convergence properties of such alternating iteration approaches. Those can be cured by formulating an appropriate variational model, which is then minimized by an alternating iteration method. In this way consistent and convergent iteration schemes can be obtained and the properties of the limits can be understood from the structure of the underlying variational problem. A key issue is that smoothness and smallness of deformations or velocities can be used as reasonable priors (via appropriate energy terms in the variational problems), which finally leads to physically reasonable reconstructions with significantly improved signal-to-noise ratio. Examples are the approaches of [41, 8, 30] to motion-corrected PET reconstruction.

Similar as in [41, 8] we are going to minimize a functional of the form

$$J(\rho, y) = \sum_{i=0}^N (D(K\rho^i, f^i) + \alpha^i R_I(\rho^i) + \beta^i R_M(y^i)) \quad (1)$$

for an appropriate transformation model. Here  $D$  is the data fidelity between estimated and measured data, and  $R_I$  respectively  $R_M$  are regularization functionals on the image respectively motion (with nonnegative regularization parameters  $\alpha^i$  and  $\beta^i$ ).

In our setup we reconstruct a sequence of nonnegative densities  $\rho^0, \dots, \rho^N$  on  $\Omega \subset \mathbb{R}^d$ , typically  $d = 2, 3$  such that

$$\rho^i(x) = \rho^0(y^i(x)) \det(\nabla y^i(x)), \quad (2)$$

with a reasonably smooth deformation field  $y^i : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . The different densities are thought of as evaluations at (ascending) time steps  $t^i$ .

The measurements  $f^i$  are noisy versions of the projected image at the  $i$ -th time step  $K(\rho^i)$ , with a stationary forward operator  $K$ , e.g. a convolution in fluorescence microscopy or versions of the x-ray transform in PET and SPECT. Motivated by the above

applications, in particular cardiac PET, we will focus on noisy data drawn from Poisson statistics, i.e.  $f^i(x)$  is interpreted as a Poisson distributed random variable with expected value  $\tau (K\rho^i)(x)$ , where  $\tau$  is a typical length of the time interval. However, extension to various other commonly used noise models, such as additive noise or additional background noise, is straight-forward in our variational setting by simply exchanging a data fidelity term related to the negative log-likelihood of the noise distribution.

The main contributions of the paper are the following:

- An appropriate modelling of the mass-conserving density transformation and its implementation in the reconstruction process, which has not been considered previously.
- A model with suitable regularization functionals for images with edges (total variation) and for physically reasonable, (weakly) one-to-one deformations (hyperelastic).
- A detailed mathematical analysis of the arising variational problem in a function space setting.
- The construction and implementation of computational methods for a setup in cardiac PET, with applications to realistic datasets.

As we shall see below, a major issue in the analysis of the model is the fact that we only deal with BV images and Sobolev deformations, whose composition is not defined in a straight-forward way and continuity with respect to weak convergence is a rather difficult issue. The analysis section is therefore split into two main parts: In the first we present some results from geometric measure theory, which allow for a generalized version of the change of variables formula. With help of this formula we are able to prove weak  $L^1$  convergence of the composition of BV functions with Sobolev mappings (Theorem 18). This convergence result is the backbone of the existence result we formulate for injective transformations (Theorem 15).

The next section lays the focus on our numerical optimization. Since we assume time discrete data, we pick a standard space discretization and perform a First-Discretize-Then-Optimize approach. Similar to [41] we end up with an alternating two step minimization: The first step is reconstruction with a fixed motion estimation, while we optimize the motion estimation with a fix reconstructed image in the second. We describe both steps in detail and show that we can perform standard EM-TV algorithms also with motion corrected reconstruction. The minimization in the motion step leads to image registration, but we present a new distance measure which is defined on the detector domain. Despite from that we follow the widely known FAIR-framework [43] for the registration.

## 2. MOTION MODELS AND VARIATIONAL FORMULATION

In the following we discuss the appropriate mathematical modelling of motion-corrected reconstruction of density images. Our goal is to obtain the reconstructed image, which is a sequence of nonnegative, integrable probability densities

$$\rho = (\rho^0, \rho^1, \dots, \rho^N) \in L_+^1(\Omega)^{N+1} \quad (3)$$

as well as the motion (deformation)

$$y = (y^1, \dots, y^N) \in H^1(\Omega)^N \quad (4)$$

as a minimizer of a suitable variational problem. We will further restrict the set of admissible deformations below and sometimes also look for densities in different spaces, e.g. as functions of bounded variation or relaxed as probability measures instead of densities. We will investigate the following variational model: Minimize

$$J(\rho, y) = \sum_{k=0}^N (D(K\rho^i, f^i) + \alpha^i |\rho^i|_{BV(\Omega)}) + \sum_{k=1}^N \beta^i S^{hyper}(y^i) \quad (5)$$

over nonnegative  $\rho$  subject to (2). Here  $K : L^1(\Omega) \rightarrow Y$  is assumed to be a compact operator and the  $f^i$  are noisy measurements at time  $t^i$ , we refer to the subsequent sections for precise definition of the BV-seminorm and  $S^{hyper}$ . This can be interpreted as a MAP estimate for a (formal) Bayesian model as we demonstrate in the following.

**2.1. Bayesian Modelling.** In order to derive a variational model we can resort to a formal Bayesian model, where for simplicity we use the above notations. For a rigorous modelling, discretized versions of the above variables or infinite-dimensional probability distributions should be used, which is beyond the scope of this paper. Bayes theorem yields the posterior probability density

$$\mathbf{P}((\rho, y)|f) \sim \mathbf{P}(f|\rho) \mathbf{P}(\rho|y) \mathbf{P}(y), \quad (6)$$

where  $\mathbf{P}(f|\rho)$  is the likelihood, which we simply assume as

$$\mathbf{P}(f|\rho) = \prod_{i=0}^N \mathbf{L}_n(K\rho^i, f^i), \quad (7)$$

where  $\mathbf{L}_n$  is the stationary noise likelihood of observing  $f$  given  $\rho$ . Note that in our model we make the natural assumption that  $\mathbf{L}_n$  does not depend on  $y$  explicitly, since the image formation and hence the involved noise is considered as an instantaneous snapshot given some density at time  $t^i$ . For data obtained by collecting information over a larger time interval (or even the full interval  $(t^{i-1}, t^i]$ , the modelling needs to be changed at this point.

The prior probability density for the image sequence given the deformations is given by

$$\mathbf{P}(\rho|y) \sim \mathbf{P}_0(\rho^0) \prod_{k=1}^N \varepsilon(\rho^i - \rho^0(y^i) \det(\nabla y^i)) \mathbf{P}_i(\rho^i), \quad (8)$$

where  $\varepsilon$  is the concentrated measure (centered at the origin) and  $\mathbf{P}_i$  is some a-priori probability for the image at time step  $i$ . Finally,  $\mathbf{P}(y)$  is a prior probability density on the deformation sequence.

In order to compute a maximum a-posteriori probability (MAP) estimator, we need to restrict to the set of images and deformations satisfying (2) due to the concentrated measure and minimize the negative log-likelihood of the remaining factors in the density. With  $D(K\rho, f) = -\log \mathbf{L}_n(K\rho^i, f^i)$  we obtain a problem of the form

$$\min_{\rho, y} \sum_{i=0}^N (D(K\rho^i, f^i) - \log \mathbf{P}_i(\rho^i)) - \log \mathbf{P}(y). \quad (9)$$

As usual the negative logarithms of the prior probabilities are related to regularization terms in standard inverse problems theory [31]. Hence, we can directly specify those using

appropriate models. Since we aim at reconstructing images with sharp edges we employ total variation regularization on the image, i.e.,

$$-\log \mathbf{P}_i(\rho^i) = \alpha^i |\rho^i|_{BV(\Omega)} \quad (10)$$

For the deformation we use a hyperelastic energy (cf. [49]), i.e.,

$$-\log \mathbf{P}(y) = \sum_{i=1}^N \beta^i S^{hyper}(y^i), \quad (11)$$

which is natural since it also regularizes the Jacobian determinant  $\det(\nabla y^i)$ . Note that hyperelastic regularization also enforces orientation preservation and thus  $\det(\nabla y) > 0$  a.e. [12]. In detail the hyperelastic energy is given by

$$S^{hyper}(y) = \int_{\Omega} \alpha_1 \text{len}(\nabla y) + \alpha_2 \text{surf}(\text{cof}(\nabla y)) + \alpha_3 \text{vol}(\det(\nabla y)) \, dx, \quad (12)$$

with the penalty functions

$$\text{len}(s) = \|s - \mathcal{I}\|_{\text{Fro}}^2, \quad \text{surf}(s) = (\max(\|s\|_{\text{Fro}}^2 - 3, 0))^2, \quad \text{vol}(s) = \begin{cases} \frac{(s-1)^4}{s^2} & \text{if } s > 0 \\ +\infty & \text{else} \end{cases}.$$

The three terms punish deviations from the identity and in volume, length and surface. This energy enforces locally one-to-one transformations, but there might be globally non injective transformations, see [53].

Having defined regularization for the density as well as for the motion field, we look for a minimizer of (5), subject to (2)

In the particular case of Poisson data, the data fidelity equals the Kullback-Leibler divergence (for  $K$  being a continuous operator from  $L_+^1(\Omega)$  to  $L_+^1(\Sigma)$ ) up to a constant independent of  $f^i$ , more precisely

$$D(K\rho^i, f^i) = \int_{\Sigma} K\rho^i - f^i \log(K\rho^i) d\sigma \quad (13)$$

### 3. ANALYSIS

This section is devoted to the analysis of the functional we derived in the previous section. Resulting from the mass-preservation condition we derived a functional with transformed images of the form  $\rho(y) \det(\nabla y)$ . The transformation theorem for integrals is a powerful tool for dealing with such transformations. Unfortunately the hyperelastic regularization we imposed on the transformation does not guarantee diffeomorphic transformations in the classical sense. Thus we describe the relaxation of classical infinitesimal calculus to Sobolev mappings, before we focus on the presentation of our analytical results.

**3.1. Preliminary Results.** In this section we generalize known definitions from the classical infinitesimal calculus to equivalence classes of functions in Lebesgue- respectively Sobolev spaces (compare for example [22, 28, 25] for a further course on this matter). Our final goal is to derive a version of the transformation theorem for integrals for non-diffeomorphic functions. Since we can not distinguish functions differing on zero sets, the classical definition of differentiability is not a feasible way, because we would like to

obtain the same result for all representatives of the equivalence class. A natural way to define a coherent function value for all representatives is via averaging, which leads to the following definition.

**Definition 1.** *Let  $\Omega$  be a domain and  $y \in L^1(\Omega)$ . Then the set of points  $x$  for which  $y_l(x)$  exists, such that*

$$\oint_{B(x,r)} |y(z) - y_l(x)| dz \rightarrow 0 \quad \text{as} \quad r \rightarrow 0 \quad (14)$$

*is called the Lebesgue set  $\mathcal{L}_y$ , while the points in  $\mathcal{L}_y$  are called Lebesgue points.*

**Remark 2.**

- *It is known that the complement of the Lebesgue points is a zero set [20, Thm. 2.19].*
- *For  $y \in W_{loc}^{1,1}$  we can define the set of Lebesgue points  $\mathcal{L}_{Dy}$  for the derivative  $Dy$  analogously.*
- *If  $y$  is a vector-valued function, we say  $x$  is a Lebesgue point, iff it is a Lebesgue point for each component function.*

Differentiability can be generalized in a similar way [2]:

**Definition 3** (Approximate differential for  $L_{loc}^1$ ). *Let  $y \in L_{loc}^1(\Omega, \mathbb{R}^m)$  and let  $x \in \mathcal{L}_y$ ; we say that  $y$  is approximately differentiable at  $x$ , iff there exists a  $d \times m$  matrix  $L$ , such that*

$$\oint_{B_r(x)} \frac{|y(z) - y_l(x) - L(z - x)|}{r} dz \rightarrow 0 \quad \text{as} \quad r \rightarrow 0. \quad (15)$$

To overcome difficulties arising from changing functions on a zero set, Nikolai Lusin imposed in his dissertation [40] the so called Lusin condition, also known as N-condition:

**Definition 4** (Lusin's condition). *A mapping  $y : \Omega \rightarrow \mathbb{R}^d$  satisfies Lusin's condition, iff:*

$$\forall E \subset \Omega, \lambda(E) = 0 \Rightarrow \lambda(y(E)) = 0 \quad (16)$$

*with  $\lambda$  denoting the Lebesgue-measure.*

Having surmounted these difficulties we need to take in account, that a non-diffeomorphic function may hit some points several times. For this problem Stefan Banach introduced the so called Banach indicatrix, which gives the number of roots to an equation [6]. This concept was generalized to discontinuous functions by Lozinski [38] and into higher dimensions by Kronrod [35] and Vitushkin in his master thesis [54]:

**Definition 5** (Banach indicatrix). *Let  $y : \mathbb{R}^d \rightarrow \mathbb{R}^m$ ,  $E \subset \mathbb{R}^d$ . The Banach indicatrix is a function*

$$N_y(\cdot, \Omega) : \mathbb{R}^m \rightarrow \mathbb{N}_0 \cup \{\infty\},$$

*which is given by*

$$N_y(z, E) := \text{card}(\{y^{-1}(z) \cap E\}). \quad (17)$$

We are now ready to present a change of variables formula under minimal assumptions, which was given by Hajlasz [28]. The central idea is, that points hit multiple times by the transformation need to be taken into account as multiplicative factor.

**Theorem 6** (Area formula). *Let  $y : \mathbb{R}^d \rightarrow \mathbb{R}^m$  be a mapping. If  $y$  is approximately differentiable almost everywhere, then  $y$  can be redefined on a zero set, such that the new  $y$  fulfills Lusin's condition. Furthermore the following statements hold for every measurable subset  $\Omega$  and positive measurable function  $u : \tilde{\Omega} \rightarrow \mathbb{R}$ :*

- (i) *The functions  $u(y) \det(\nabla y)$  and  $u(z)N_y(z, \Omega)$  are measurable.*
- (ii) *If moreover  $u \geq 0$  then*

$$\int_{\Omega} u(y(x)) \det(\nabla y(x)) dx = \int_{\mathbb{R}^m} u(z) N_y(z, \Omega) dz. \quad (18)$$

- (iii) *If one of the functions  $u(y) \det(\nabla y)$  and  $u(z)N_y(z, E)$  is integrable then so is the other and the formula (18) holds.*

*Additionally we have*

$$\int_{\Omega} u(x) \det(\nabla y(x)) dx = \int_{\mathbb{R}^d} \sum_{w \in (y^{-1}(z) \cap \Omega)} u(w) dz. \quad (19)$$

*Proof.* See [28, Theorem 2] for the first part of the theorem. The second part is given in [25, Chapter 1.2 Theorem 2] for Lipschitz mappings and can be generalized by following the proof by Hajlasz [28, Theorem 2].  $\square$

As we see, the generalization of the transformation theorem can take into account, that points might be hit several times. Consequently a mapping with no additional injectivity restriction might violate the mass-preservation condition.

Furthermore, this subject is strongly linked to the topological degree, which was introduced by Brouwer in 1911 [10] and generalized to Sobolev mappings by Giaquinta et al. [24]. We will not focus on this topic, but we will use some of these results in the remaining part of this thesis. We start by giving the generalized definition of the topological degree, see also [24]

**Definition 7** (Topological degree). *Let  $\Omega \subset \mathbb{R}^d$  be an open set and  $y$  an almost everywhere approximately differentiable map with Jacobian  $Dy$  the degree of  $y$  is defined as*

$$\deg(y, \Omega, z) := \sum_{x \in y^{-1}(z)} \text{sgn}(\det(Dy(x))) \quad (20)$$

**Remark 8.** *The topological degree is strongly related to the Banach indicatrix, since it coincides with the Banach indicatrix for certain mappings. As a direct consequence of [23, Chapter 1, Proposition 2] we obtain for orientation preserving mappings:*

$$\deg(y, \Omega, \cdot) = N_y(\cdot, \Omega) \quad a.e. \quad (21)$$

An interesting property of the topological degree is that it is completely determined on the boundary for sufficiently regular functions. This is phrased in the following proposition.

**Proposition 9.** *Let  $\Omega$  be a bounded Lipschitz domain in  $\mathbb{R}^d$  and let  $y_1, y_2$  be mappings in  $W^{1,d-1}(\Omega, \mathbb{R}^d)$  with  $\text{cof}(\nabla y_i) \in L^{\frac{d}{d-1}}$ . Suppose that  $y_1 = y_2$  text  $\partial\Omega$  in the sense of  $W^{1,d-1}$  traces. Then*

$$\deg(y_1, \Omega, z) = \deg(y_2, \Omega, z)$$

*Proof.* See [23], Chapter 2, Proposition 1. □

To conclude this brief summary, we mention that there is indeed a specification to injective Sobolev functions. This leads to the field of weak diffeomorphisms [21]. With Theorem 10 we will only present a compactness result, which is central for our analysis. A short introduction is given in the appendix. A weak diffeomorphism  $y$  can be expressed as the limit of a sequence  $y_n$  of orientation preserving  $C^1$  diffeomorphisms [21]. Furthermore closeness and compactness results for this class can be stated ([23, Chapter 5, Theorems 3 and 4]). This relies heavily on weak convergence in the space of weak diffeomorphisms  $\tilde{\text{dif}}^{p,q}$ . Thus we present a similar result published by Henao and Mora-Corral [29] instead, which follows directly from [29, Proposition 2 and Theorem 2].

**Theorem 10** (Injectivity as closed constraint). *For each  $j \in \mathbb{N}$  let  $y_j, u : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$  be a.e. approximately differentiable. Assume furthermore, that*

$$y_j \in W^{1,p}(\Omega, \mathbb{R}^m) \quad p \geq d-1 \quad \det(Dy_j) \in L^1(\Omega) \quad (22)$$

as well as

$$\text{cof}(Dy_j) \in L^q(\Omega) \quad q \geq \frac{p}{p-1} \quad \sup_{j \in \mathbb{N}} \|\text{cof}(Dy_j)\|_1 < \infty. \quad (23)$$

Suppose that there exists  $\theta \in L^1(\Omega)$  such that  $\theta > 0$  a.e. and

$$y_j \rightarrow y \quad \det(Dy_j) \rightharpoonup \theta \text{ in } L^1(\Omega) \quad (24)$$

as  $j \rightarrow \infty$ . Assume that for each  $j \in \mathbb{N}$  the function  $y_j$  we have  $N_{y_j} \leq 1$  a.e. with  $\det(Dy_j) > 0$  a.e.. Then

- (i)  $\theta = |\det(Dy)|$  a.e.,
- (ii)  $N_y \leq 1$  a.e..

**3.2. Regularization Functionals.** In this section we shortly present the properties of the regularization functionals we use in our reconstruction framework. Both regularizations guarantee stronger convergence properties additionally to the compactness of sub-level sets. We do not focus on general possibilities for choosing these energies, but rather present the specific choices for images (total variation) and transformation (hyperelastic).

Total variation regularization was first introduced for image denoising in [48], now known as the ROF-model. Recently TV regularization has been applied to different tasks in imaging. In fact there are many different equivalent ways to define total variation and BV-functions (compare [5, Chapter 10, Definition 10.1.1]), but we focus on the common definition in image processing (cf. [13]):

**Definition 11** (BV seminorm and functions of bounded variation). *Let  $u : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ . Then the BV-seminorm is given by*

$$|u|_{BV(\Omega)} := \sup_{g \in C_0^\infty(\Omega; \mathbb{R}^d), \|g\|_\infty \leq 1} \int_{\Omega} u \nabla \cdot g dx. \quad (25)$$

Consequently we define the space of functions with bounded variation  $BV(\Omega)$  by

$$BV(\Omega) := \{u \in L^1(\Omega) \mid |u|_{BV(\Omega)} < \infty\} \quad \|u\|_{BV} := \|u\|_1 + |u|_{BV(\Omega)} \quad (26)$$



The BV-seminorm is lower semicontinuous in the  $L^1_{loc}$  topology [2, Remark 3.5]. [2, Remark 3.12], thus one can define a weak-star convergence.

Note that due to compactness of the embedding operator the weak star convergence in  $BV$  guarantees strong convergence in  $L^1$ . As we will see later, this will be useful to prove convergence properties of compositions of functions. Furthermore we can approximate any  $BV$ -function by a sequence of smooth functions:

**Theorem 12** (Approximation by smooth functions). *Let  $u \in L^1(\Omega; \mathbb{R}^m)$ . Then  $u \in BV(\Omega, \mathbb{R}^m)$ , if and only if there exists a sequence  $u_n$  in  $C^\infty(\Omega, \mathbb{R}^m)$  converging to  $u$  in  $L^1(\Omega, \mathbb{R}^m)$  and there exists a constant  $L$  satisfying*

$$\lim_{n \rightarrow \infty} \int_{\Omega} |\nabla u_n| dx < \infty \quad (27)$$

*Proof.* [2, Thm. 3.9]. □

**3.2.1. Hyperelastic Regularization.** For the hyperelastic regularization energy, we presented earlier, Ruthotto [49] used the following set of admissible transformations

$$\mathcal{A} := \{y \in \mathcal{A}_0 : \left| \int_{\Omega} y(x) dx \right| \leq \text{vol}(\Omega)(M + \text{diam}(\Omega))\}, \quad (28)$$

where  $\Omega$  is bounded by  $M$ , and  $\mathcal{A}_0$  is defined by:

$$\begin{aligned} \mathcal{A}_0 := \{y \in W^{1,2} : \text{cof}(\nabla y) \in L^4(\Omega, \mathbb{R}^{3 \times 3}) \\ \det(\nabla y) \in L^2(\Omega, \mathbb{R}), \det(\nabla y) > 0 \text{ a.e.}\} \end{aligned}$$

**Remark 13.** *A transformation  $y \in \mathcal{A}_0$  is call  $y$  an **admissible transformation**. Note that all admissible transformations fulfill the conditions on the cofactor in Proposition 9 and Theorem 10.*

As we see the admissible transformations have strict positive Jacobians a.e. and are thus locally invertible. Note that this deduction is not trivial: Since by the Sobolev embedding theorem [18, Chapter 5.6.4, Thm 6] an admissible function does not need to have a continuous version, we cannot use the implicit function theorem to deduce the existence of a local inversion. Even worse, the standard theory on (local) invertibility of Sobolev mappings is focussed on mappings in  $W^{1,d}$  (see for example [19, 34]), so this theory would only be applicable for  $d = 2$ . However, we can use some results from the theory of Cartesian currents we mentioned briefly earlier and use the fact, that:

$$\mathcal{A} \subset \mathcal{A}_{d-1,d-1}(\Omega, \mathbb{R}^m) \quad (29)$$

with  $\mathcal{A}_{d-1,d-1}(\Omega, \mathbb{R}^m)$  as being defined in the Appendix. Then a result from Müller [45] yields the closedness of the graph [24] of the transformation  $y$  and thus  $y$  is a weak local diffeomorphism as defined by Giaquinta et al. [26]. We will not elaborate on this further, because we used this argumentation only to demonstrate that we can expect to have local invertibility and that this property is not directly guaranteed by the positivity of the Jacobian determinant.

We conclude this short course on hyperelastic regularization by stating the convergence properties, shown in [49, Chapter 3, Theorem 4]:

**Theorem 14** (Convergence properties of admissible transformations). *Let  $\Omega$  be a domain with a  $C^1$  boundary and  $y_k, y \in \mathcal{A}_0$  be admissible transformations. Then convergence of*

$$y_k \rightharpoonup y \text{ in } W^{1,2}(\Omega, \mathbb{R}^3), \quad \text{cof}(\nabla y_k) \rightharpoonup H \text{ in } L^4(\Omega, \mathbb{R}^{3 \times 3}), \quad \det(\nabla y_k) \rightharpoonup v \text{ in } L^2(\Omega, \mathbb{R})$$

*implies  $H = \text{cof}(\nabla y)$  and  $v = \det(\nabla y)$ .*

**3.3. Existence of a Minimizer.** In this section we establish the following existence result for a compact subset  $\mathcal{K}$  of the injective admissible transformations:

**Theorem 15** (Existence of a minimizer in motion-corrected reconstruction). *Let our assumptions (A1)-(A3) hold, furthermore we assume that we have*

$$J(1, (Id)^N) = \sum_{i=0}^N (D(K\rho^0, f^i) + \alpha_i |\rho^0|_{BV(\Omega)}) + \sum_{i=1}^N \beta_i S^{hyper}(Id) < \infty, \quad (30)$$

*Then the functional (5)*

*with a  $L^1$ -continuous distance term fulfilling coercivity property (47) has at least one minimizer*

*$(\hat{\rho}, \hat{y}) \in BV(\Omega) \times (\mathcal{K})^N$ , where  $\mathcal{K}$  is a closed subset of  $\mathcal{A} \cap \mathcal{I}$ . Particularly this holds for the Kullback-Leibler divergence as distance term.*

Injectivity is crucial to enforce the mass-preservation condition. Despite the fact that our proof can be extended to transformations with  $N_y(\cdot, \Omega)$  bounded in  $L^\infty$  we shortly give two closed subsets of the injective transformations:

**Remark 16.** • *The set  $\mathcal{A} \cap \mathcal{I}$  is closed [29].*

- *The set  $\mathcal{B}_v := \{y \in \mathcal{A} : y|_{\partial\Omega} = v\}$  is a closed subset of  $\mathcal{A} \cap \mathcal{I}$  for an injective boundary value  $v$ , where the equality is understood in the sense of  $H^1$  traces.*

Crucial for the proof is controlling sequences of functions after transformations  $\rho_k(y_k) \det(\nabla y_k)$  with  $\rho_k, y_k$  being weak-\* convergent sequences resulting from coercivity properties. Before actually presenting our central theorem on this convergence, we define weak-\* convergence for the transformation as follows (compare [50]).

**Definition 17** (Weak-\* Convergence in  $\mathcal{A}$ ). *Let  $y_k, y \in \mathcal{A}$ . We say  $y_k \rightharpoonup^* y$  in  $\mathcal{A}$ , iff*

$$\begin{aligned} y_k &\rightharpoonup^* y && \text{in } W^{1,2}(\Omega, \mathbb{R}^3) \\ \text{cof}(\nabla y_k) &\rightharpoonup^* \text{cof}(\nabla y) && \text{in } L^4(\Omega, \mathbb{R}^{3 \times 3}) \\ \det(\nabla y_k) &\rightharpoonup^* \det(\nabla y) && \text{in } L^2(\Omega, \mathbb{R}) \end{aligned}$$

Now we can deduce convergence properties of composed weak-\* convergent sequences.

**Theorem 18.** *Let  $\rho_k^0 \rightharpoonup^* \rho^0$  in  $BV(\Omega_0)$ ,  $y_k \rightharpoonup y$  in  $H^1(\Omega)$  and  $\det(\nabla y_k) \rightharpoonup \det(\nabla y)$  in  $L^2(\Omega)$ . Assume additionally, that*

$$\sup_k \|N_{y_k}(\cdot, \Omega)\|_\infty \leq C \in \mathbb{R}. \quad (31)$$

*Then we obtain  $\rho_k^0(y_k) \det(\nabla y_k) \rightharpoonup \rho^0(y) \det(\nabla y)$  in  $L^1(\Omega)$ .*

*Proof.* For the ease of presentation we use the following abbreviation:

$$\rho_k := \rho_k^0 \quad \rho := \rho^0 \quad d_k := \det(\nabla y_k)$$

For any fixed  $\varphi \in (L^1)^* = L^\infty$  we show:

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} \int_{\Omega} (\rho_k(y_k) d_k - \rho(y) d) \varphi dx \\ &= \lim_{k \rightarrow \infty} \int_{\Omega} (\rho_k(y_k) d_k - \rho(y_k) d_k + \rho(y_k) d_k - \rho(y) d) \varphi dx \\ &= \lim_{k \rightarrow \infty} \int_{\Omega} (\rho_k(y_k) d_k - \rho(y_k) d_k) \varphi dx + \lim_{k \rightarrow \infty} \int_{\Omega} (\rho(y_k) d_k - \rho(y) d) \varphi dx \end{aligned}$$

We examine both summands separately and show that each of them converges to zero. We recall that weak star convergence in  $BV$  implies strong convergence in  $L^1$  [13]. Now the first term is treated in a straightforward way by the area formula and (31)

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\Omega} |\rho_k(y_k) d_k - \rho(y_k) d_k| dx &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} |\rho_k - \rho| N_{y_k}(x, \Omega) dx \\ &\leq \lim_{k \rightarrow \infty} C \int_{\mathbb{R}^n} |\rho_k - \rho| dx \end{aligned}$$

Since  $\text{supp}(\rho_k - \rho) \subseteq \Omega_0$  and  $\rho_k \rightarrow \rho$  in  $L^1(\Omega_0)$  implies weak convergence, we obtain

$$\lim_{k \rightarrow \infty} \int_{\Omega} (\rho_k(y_k) d_k - \rho(y_k) d_k) \varphi dx = 0 \quad \forall \varphi. \quad (32)$$

The second term needs further care: According to [2, Thm. 3.9]. we find a sequence of functions  $(\xi^n)_n \subset C^\infty(\Omega_0)$  with  $\lim_{n \rightarrow \infty} \|\xi^n - \rho\|_1 = 0$ .

Let  $\epsilon > 0$ . Thus we can pick a fix  $N$ , such that:

$$\int_{\Omega_0} |\xi^N - \rho| dx \leq \frac{\epsilon}{4 \|\varphi\|_\infty C}. \quad (33)$$

We now expand the first term with said  $\xi^N$  and obtain:

$$\begin{aligned}
& \int_{\Omega} (\rho(y_k)d_k - \xi^N(y_k)d_k + \xi^N(y_k)d_k - \xi^N(y)d_k \\
& + \xi^N(y)d_k - \xi^N(y)d + \xi^N(y)d - \rho(y)d)\varphi dx \\
& = \int_{\Omega} (\rho(y_k)d_k - \xi^N(y_k)d_k)\varphi dx + \int_{\Omega} (\xi^N(y_k)d_k - \xi^N(y)d_k)\varphi dx \\
& + \int_{\Omega} (\xi^N(y)d_k - \xi^N(y)d)\varphi dx + \int_{\Omega} (\xi^N(y)d - \rho(y)d)\varphi dx,
\end{aligned}$$

We examine each term separately: We start with applying Hölder's inequality to the first one and obtain

$$\begin{aligned}
\int_{\Omega} (\rho(y_k)d_k - \xi^N(y_k)d_k)\varphi dx & \leq \int_{\Omega} |\rho(y_k)d_k - \xi^N(y_k)d_k| dx \|\varphi\|_{\infty} \\
& = \int_{\mathbb{R}^n} |\rho - \xi^N| N_{y_k}(z, \Omega) dz \|\varphi\|_{\infty} \\
& \leq C \|\rho - \xi^N\|_1 \|\varphi\|_{\infty},
\end{aligned}$$

which yields together with (33)

$$\int_{\Omega} (\rho(y_k)d_k - \xi^N(y_k)d_k)\varphi dx \leq \frac{\epsilon}{4}. \quad (34)$$

In the second term we start in a straightforward way with

$$\int_{\Omega} |\xi^N(y_k)d_k| dx \leq \|\xi^N(y_k)\|_2 \|d_k\|_2. \quad (35)$$

By using that  $\xi^N(y_k)$  is bounded,  $\Omega$  compact and  $y_k$  is weakly convergent, it directly follows, that  $\|\xi^N(y_k)\|_2 < \infty$  and  $\|d_k\|_2$  is bounded by some  $E \in \mathbb{R}$  (Banach-Steinhaus, see e.g. [1, Chapter 5, Theorem 3]). Thus (35) is finite. Aswell we can directly deduce

$$\int_{\Omega} |\xi^N(y)d_k| dx \leq \|\xi^N(y)\|_2 \|d_k\|_2 \leq \infty, \quad (36)$$

and hence,

$$\begin{aligned}
\int_{\Omega} (\xi^N(y_k)d_k - \xi^N(y)d_k)\varphi dx & \leq \|(\xi^N(y_k) - \xi^N(y))\varphi\|_2 \|d_k\|_2 \\
& \leq \|(\xi^N(y_k) - \xi^N(y))\varphi\|_2 E
\end{aligned}$$

Note that  $\xi^N$  is in  $C^\infty$  and therefore  $(\xi^N)^2$  is Lipschitz-continuous with some constant  $L$ , since  $\Omega$  is compact. We obtain

$$\int_{\Omega} ((\xi^N(y_k) - \xi^N(y))\varphi)^2 dx \leq (\|\varphi\|_\infty)^2 L \|y_k - y\|_2. \quad (37)$$

By the Rellich-Kondrachov compactness theorem [18, Chapter 5.7, Theorem 1]  $y_k$  converges strongly to  $y$  in  $L^2$  and so we find  $K_2 \in \mathbb{N}$ , such that

$$\|y_k - y\| < \frac{\epsilon^2}{16 (\|\varphi\|_\infty)^2 L E^2}. \quad \forall k \geq K_2 \quad (38)$$

This implies for each  $k \geq K_2$

$$\int_{\Omega} (\xi^N(y_k)d - \xi^N(y)d)\varphi dx < \frac{\epsilon}{4}. \quad (39)$$

Now let us consider the third term. Since  $\xi^N(y_k)$  is bounded it follows from the weak convergence of the determinants, that there exists  $K_3 \in \mathbb{N}$ , such that for every  $k \geq K_3$

$$\int_{\Omega} (\xi^N(y)d_k - \xi^N(y)d)\varphi dx \leq \frac{\epsilon}{4}. \quad (40)$$

Note that the compactness of  $\Omega$  ensures that  $\varphi\xi^N(y) \in L^2$ . For the last term we can proceed as for the first one, with  $K_1$  as above, and deduce

$$\int_{\Omega} (\xi^N(y)d - \rho(y)d)\varphi dx \leq \frac{\epsilon}{4} \quad (41)$$

for any  $k \geq K_1$ .

By combining (34), (39), (40) and (41), we obtain for every  $k \geq K := \max\{K_1, K_2, K_3\}$

$$\int_{\Omega} (\rho(y_k)d_k - \rho(y)d)\varphi dx \leq \epsilon \quad \forall \varphi$$

and thus

$$\lim_{k \rightarrow \infty} \int_{\Omega} (\rho_k(y_k)d_k - \rho_k(y)d)\varphi dx = 0. \quad (42)$$

The assertion follows by combining (32) and (42).  $\square$

As we see, the boundedness of the Banach-Indicatrix is substantial to control sequences of images after transformations. In order to guarantee the convergence properties given by Theorem 18, we restrict our analysis to injective transformations. Additionally this has the effect that as a consequence of the area formula the mass-preservation condition is not violated.

Now we can give continuity properties for a wide range of distance measures including the Kullback-Leibler divergence:

**Lemma 19.** *Let the assumptions from Theorem 18 be fulfilled. Then the distance part of our functional  $J$ , defined by*

$$D(\rho_0, y) = \sum_{i=0}^N \int_{\Sigma} g(K\rho^0(y^i) \det(\nabla y^i), f^i) d\sigma \quad (43)$$

*with a  $L^1$ -continuous integrand function  $g$  is lower semicontinuous with respect to weak-star convergence in  $BV(\Omega_0)$ , weak convergence in  $W^{1,2}(\Omega_i)$  for the transformations and weak convergence in  $L^2(\Omega_i)$  for the determinants.*

*Proof.* Let  $y_k \rightharpoonup y$  in  $W^{1,2}(\Omega)$ ,  $\det(\nabla y_k) \rightharpoonup \det(\nabla y)$  in  $L^2$ ,  $\rho_k^0 \rightharpoonup \rho^0$  in  $BV(\Omega)$ . Since we denote  $y = (y^1, \dots, y^N)$  as the collection of all transformations we understand the weak convergence componentwise. From Theorem 18 we obtain for any fixed  $1 \leq i \leq N$

$$\rho_k^0(y_k^i) \det(\nabla y_k^i) \rightharpoonup \rho^0(y^i) \det(\nabla y^i).$$

Since  $K$  is a compact operator,  $K$  is completely continuous, which gives us:

$$K(\rho_k^0(y_k^i) \det(\nabla y_k^i)) \rightarrow K(\rho^0(y^i) \det(\nabla y^i)) \quad \text{in } L^1(\Sigma).$$

Therefore we can follow the proof of Lemma 3.4 (iii) from [47] for each summand and obtain lower semicontinuity of the Kullback-Leibler data fidelity term with the following reasoning:

Since  $K(\rho_k^0(y_k^i) \det(\nabla y_k^i)) \rightarrow K(\rho^0(y^i) \det(\nabla y^i))$ , we have convergence almost everywhere. Thus we can deduce

$$g(K\rho_k^0(y_k^i) \det(\nabla y_k^i)) \rightarrow g(K\rho^0(y^i) \det(\nabla y^i)) \quad \text{in } L^1(\Sigma).$$

Now we can apply Fatou's Lemma [25, Chapter 1, Theorem 2 (iii)] and obtain:

$$\int_{\Sigma} g(K\rho^0(y^i) \det(\nabla y^i)) d\sigma \leq \liminf_{k \rightarrow \infty} \int_{\Sigma} g(K\rho_k^0(y_k^i) \det(\nabla y_k^i)) d\sigma$$

Having shown lower semicontinuity for an arbitrary summand with fixed  $i$ , the assertion follows directly.  $\square$

We have now stated lower semicontinuity results for a wide range of distance terms, including the Kullback-Leibler data fidelity. We now turn our focus to the TV-regularization. By setting  $\alpha^k = 0$  for any  $1 \leq k \leq N$  in (5), this would follow directly by the properties of the BV-seminorm we mentioned earlier. However to formulate an existence result for  $\alpha^k \geq 0$ , we give the following lemma, which follows with a proof as in [13]:

**Lemma 20.** *Let  $\rho_k \rightharpoonup \rho$  in  $L^1(\Omega)$ . Then*

$$|\rho|_{BV(\Omega)} \leq \liminf_{k \rightarrow \infty} |\rho_k|_{BV(\Omega)}. \quad (44)$$

We verified the first condition for the existence of a minimizer and turn our focus coercivity.

**Lemma 21** ( $\rho$ -Coercivity for the Kullback-Leibler divergence). *Let our assumptions (28) for our transformation and (A1)-(A3) for the operator be fulfilled. Then*

$$J^1(\rho^0, y) := \sum_i \left[ \int_{\Sigma} K(\rho^0(y^i \det(\nabla y^i))) - f^i \log(K(\rho^0(y^i \det(\nabla y^i)))) d\sigma \right. \\ \left. + \alpha^i |\rho^0(y^i) \det(\nabla y^i)|_{BV(\Omega)} \right]$$

is coercive with respect to the variable  $\rho^0$ .

*Proof.* We begin by observing:

$$\log(x) \leq \frac{1}{a}x + \log(a) \quad \forall x > 0$$

for any fixed  $a > 0$ . We add the constant  $f \log(f) - f$ , such that each summand is non-negative. Note that  $y^0 = Id$  and therefore we can bind the Kullback-Leibler divergence from below by:

$$\begin{aligned} & \sum_{i=0}^N \int_{\Sigma} K(\rho^0(y^i) \det(\nabla y^i)) - f^i \log(K(\rho^0(y^i) \det(\nabla y^i))) + f^i \log(f^i) - f^i d\sigma \\ & \geq \int_{\Sigma} K(\rho^0(y^0) \det(\nabla y^0)) - f^0 \log(K(\rho^0(y^0) \det(\nabla y^0))) + f^0 \log(f^0) - f^0 d\sigma \\ & = \int_{\Sigma} K\rho^0 - f^0 \log(K\rho^0) + f^0 \log(f^0) - f^0 d\sigma \\ & \geq \int_{\Sigma} K\rho^0 - f^0 \left( \frac{1}{f^0 + 1} K\rho^0 + \log(f^0 + 1) \right) + f^0 \log(f^0) - f^0 d\sigma \\ & \geq \int_{\Sigma} \frac{1}{f^0 + 1} K\rho^0 d\sigma - \underbrace{\int_{\Sigma} f^0(f^0 + 1) + f^0 \log(f^0) - f^0 d\sigma}_{:= c_2 \in \mathbb{R}} \\ & = \int_{\Omega} K^* \frac{1}{f^0 + 1} \rho^0 dx + c_2 \geq \int_{\Omega} c_1 \rho^0 dx + c_2 dx \\ & = c_1 \|\rho^0\|_1 + c_2 \end{aligned}$$

Therefore we can conclude

$$J_1(\rho^0, y) \geq c_1 \|\rho^0\|_1 + \alpha^0 |\rho^0|_{BV(\Omega)} \geq \min\{c_1, \alpha^0\} (\|\rho^0\|_1 + |\rho^0|_{BV(\Omega)}). \quad (45)$$

□

**Remark 22.** *The lemma holds not only for the Kullback-Leibler divergence, but for any distance of the form*

$$D(K\rho^i, f^i) = \int_{\Sigma} g^i(K\rho^i) d\sigma \quad (46)$$

with  $g^i$  convex, which is bounded below and satisfies

$$g^0(K\rho^0) \geq c_1 K\rho^0 + c_2 \quad (47)$$

with constants  $c_1 > 0$  and  $c_2 \in \mathbb{R}$ , only dependent of  $f$ .

We have proved that the first part of our functional is coercive in  $\rho_0$ , so it remains to be shown, that the hyperelastic regularization is coercive with respect to  $y$ . Fortunately this has already been done in [49, Chapter 2, Lemma 1]. Combining these results, we can finally prove Theorem 15:

*Proof.* (of Theorem 15) The coercivity properties of distance (47), TV and hyperelastic regularization ensure the existence of a bounded level set. Compactness is then granted by the Banach-Alaoglu Theorem. Thus a minimizing sequence  $(\rho_k, y_k)$  has a convergent subsequence with limit  $(\rho, y)$ , while the convergence is understood component wise. By the lower semicontinuity properties of distance (Lemma 19), TV (Lemma 20) and hyperelastic regularization we obtain:

$$J(\rho, y) \leq \liminf J(\rho_k, y_k) = \inf J$$

Finally as a result of the compactness of  $\mathcal{K}$  we have  $(\rho, y) \in BV(\Omega) \times \mathcal{K}^N$ .  $\square$

**3.4. Convergence of the Regularization Method.** Having stated existence results for the variational problem of motion-corrected reconstruction we can show with analogous weak compactness and lower semicontinuity arguments that the minimization of our functional (5) can be understood as a (nonlinear) regularization method, i.e. there exist appropriate limits as noise and regularization parameter tend to zero [17].

**Theorem 23.** *Let  $(f_k^i)_k$  be a sequence of noisy data with*

$$\lim_k f_k^i = f_*^i, \quad (48)$$

where  $f^*$  is the exact data for an image  $\rho_*$  and transformations  $y_*$ , such that:

$$(\rho_*, y_*) = \min_{\rho, y} \sum_{i=0}^N D(K(\rho(y^i) \det(\nabla y^i)), f_*^i) \quad (49)$$

for a non-negative distance  $D$ , which is lower semicontinuous in both arguments and fulfilling (47). Furthermore, we define a sequence of functionals  $J_k$  by:

$$J_k = \sum_{i=0}^N (D(K(\rho^0(y^i) \det(\nabla y^i)), f_k^i) + \alpha_k^i |\rho^i|_{BV(\Omega)}) + \sum_{i=1}^N \beta_k^i S^{\text{hyper}}(y^i) \quad (50)$$

Then for  $\alpha_k \rightarrow 0$  and  $\beta_k \rightarrow 0$  with

$$\frac{\sum_{i=0}^N D(f_*^i, f_k^i)}{\min_i \{a_k^i, \beta_k^i\}} \rightarrow 0 \quad \frac{\max_i \{a_k^i, \beta_k^i\}}{\min_i \{a_k^i, \beta_k^i\}} \leq C \in \mathbb{R} \forall k \quad (51)$$

the sequence  $(\hat{\rho}_k, \hat{y}_k)_k$ , with  $(\hat{\rho}_k, \hat{y}_k)$  being minimizers of  $J_k$  has a convergent subsequence and the limit  $(\hat{\rho}, \hat{y})$  fulfills:

$$\sum_i D(\hat{\rho}(\hat{y}^i) \det(\nabla \hat{y}^i), f_*^i) = \sum_i D(\rho_*(y_*^i) \det(\nabla y_*^i), f_*^i) \quad (52)$$



Note that the estimation (52) only holds for the composition and not for the components. Since different transformations can lead to the same transformed image we can in general not expect to derive convergence for both components.

**Remark 24.** *In order to deal with the non uniqueness of the solution yielded by Theorem 23 we assume additionally*

- (1)  $\lim_k \frac{\sum_{i=0}^N D(K\rho_*^i, f_k^i)}{\min_i \{\alpha_k^i, \beta_k^i\}} = 0$  ,  
 (2) For all  $\alpha_k^i, \beta_k^i$  there exist  $\tilde{\alpha}^i = \lim_k \frac{\alpha^i}{\min_i \{\alpha_k^i, \beta_k^i\}}$ , resp.  $\tilde{\beta}^i = \lim_k \frac{\beta^i}{\min_i \{\alpha_k^i, \beta_k^i\}}$ ,

then we can deduce

$$\begin{aligned} & \sum_{i=0}^N \tilde{\alpha}^i |\hat{\rho}^i|_{BV(\Omega)} + \sum_{i=1}^N \tilde{\beta}^i S^{hyper}(\hat{y}^i) \\ \leq & \liminf_k \sum_{i=0}^N \tilde{\alpha}^i |\hat{\rho}_k^i|_{BV(\Omega)} + \sum_{i=1}^N \tilde{\beta}^i S^{hyper}(\hat{y}_k^i) \\ \leq & \lim_k \left( \frac{\sum_{i=0}^N D(K\rho_*^i, f_k^i)}{\min_i \{\alpha_k^i, \beta_k^i\}} + \sum_{i=0}^N \tilde{\alpha}^i |\rho_*^i|_{BV(\Omega)} + \sum_{i=1}^N \tilde{\beta}^i S^{hyper}(y_*^i) \right) \\ = & \sum_{i=0}^N \tilde{\alpha}^i |\rho_*^i|_{BV(\Omega)} + \sum_{i=1}^N \tilde{\beta}^i S^{hyper}(y_*^i). \end{aligned}$$

By using this deduction we can show that  $(\hat{\rho}, \hat{y})$  is a solution, which minimizes

$$\sum_{i=0}^N \tilde{\alpha}^i |\cdot|_{BV(\Omega)} + \sum_{i=1}^N \tilde{\beta}^i S^{hyper}(\cdot) \quad \forall \quad (\rho, y) \quad \text{with} \quad \sum_{i=0}^N D(K\rho^i, f_*^i) = 0.$$

This solution can be viewed analogously to the best-approximate solution in the sense of Engl et al. [17, Definition 2.1].

#### 4. NUMERICAL SOLUTION

In this section we describe the numerical framework we use to solve the motion-corrected reconstruction problem. We restrict the presentation of the numerical framework to the Kullback-Leibler divergence as distance measure; nevertheless the extension to other distances with given TV-regularized reconstruction algorithms is straightforward.

We aim to perform a First-Discretize-then-Optimize approach combined with an alternating minimization strategy. For this we only need a discretization in the space-domain, since we assume to have time-discretized data. For the time-discretization we assume, that the discrete data  $(f_i)_i$  is gated: Therefore we define time nodes  $(t_i)_i$  such that each node  $t_i$  represents a stage of motion (f.e. cardiac or respiratoric gate). Additionally we impose regularization only on the reference configuration  $\rho_0$  and not on the transformed versions of  $\rho_0$ , so by setting  $\alpha^i = 0$  for  $i > 0$  we obtain the following functional to be minimized

$$J(\rho_0, y) = \sum_i \int_{\Sigma} K(\rho^0(y^i) \det(\nabla y^i)) - f^i \log(K(\rho^0(y^i(\det(\nabla y^i)))))) d\sigma \\ + |\rho^0|_{BV(\Omega)} + \sum_i S^{hyper}(y^i).$$

Out discretization is straightforward: We define pairwise disjoint pixel  $B_i$ , such that

$$\Omega = \bigcup_i B_i. \quad (53)$$

With this discretization at hand we can put our problem in a discrete framework and start to minimize our functional. Given an initial value  $\rho_0^0$  our algorithm reads as follows:

$$\begin{cases} 1. \text{ Motion step:} & y_{k+1} \in \arg \min_y \{J(\rho_k^0, y)\} \\ 2. \text{ Reconstruction step:} & \rho_{k+1}^0 \in \arg \min_{\rho^0} \{J(\rho^0, y_{k+1})\} \end{cases}.$$

While the reconstruction step can be realized via motion corrected EM-TV algorithms (see e.g. [11]), the motion step needs some caretaking. In the next sections we want to outline both implementations briefly.

**4.1. Reconstruction-Step: Motion-Corrected EM-TV.** In this section we want to show that we can apply standard EM-TV algorithms to motion-corrected reconstructions. For ease of presentation we assume that the transformations  $y^i$  are global invertible, as e.g. in the case of Dirichlet boundary conditions (Proposition 9). The central idea for EM-TV is to alternate an EM step with given discrete projection operator (so called system matrix) with a TV denoising step [51]. For a classical reconstruction problem

$$\min_{\rho} \int_{\Sigma} K\rho - f \log(K\rho) + \alpha |\rho|_{BV(\Omega)} \quad (54)$$

the FB-EMTV algorithm is given by alternation of a reconstruction (EM step) with a denoising (TV) step:

$$\begin{cases} u_{k+\frac{1}{2}} = \frac{u_k}{K^*1_{\Sigma}} K^* \left( \frac{f}{Ku_k} \right) \\ u_{k+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^*1_{\Sigma} (u - u_{k+\frac{1}{2}})^2}{u_k} dx + \alpha |u|_{BV(\Omega)} \right\} \end{cases}$$

We will show how to modify the distance term, so that it fits the structure described in (54). Since the transformation is a linear operator acting on the image, we can write our transformation model (2) equivalently as

$$\rho^i = T_{y^i}^{mp} \rho_0. \quad (55)$$

Projecting the transformed reference configuration  $\rho^i$  into the measurement domain  $\Sigma$  is then given by

$$A^0 \rho^i = A^0 \left( T_{y^i}^{mp} \rho^0 \right). \quad (56)$$

Now using the associativity property leads to

$$A^0 \rho^i = \underbrace{\left( A^0 T_{y^i}^{mp} \right)}_{=: A^i} \rho^0, \quad (57)$$

which gives us a motion-corrected projection operator  $A^i$ , acting on  $\rho^0$  only. Now we can formulate the motion-corrected reconstruction problem as

$$\min_{\rho} \int_{\Sigma} \underbrace{\begin{pmatrix} A^0 \\ A^1 \\ \dots \\ A^N \end{pmatrix}}_{:=K} \rho^0 - \underbrace{\begin{pmatrix} f^0 \\ f^1 \\ \dots \\ f^N \end{pmatrix}}_{:=f} \log(A \rho^0) + \alpha |\rho^0|_{BV(\Omega)} \quad (58)$$

After this tuning our problem is suited for the the EM-TV algorithm [51] with  $u := \rho^0$ .

**4.2. Motion-Step: Interpretation as Registration.** In the Motion-Step we need to minimize

$$\begin{aligned} & \sum_i \int_{\Sigma} K(\rho_0(y^i) \det(\nabla y^i)) - f_i \log(K(\rho_0(y^i) \det(\nabla y^i))) d\sigma \\ & + |\rho_0|_{BV(\Omega)} + \sum_i S^{hyper}(y^i) \end{aligned}$$

with respect to our set of transformations  $(y^i)_i$ . Note that the problem decouples, so we obtain a minimum of the sum, by minimizing each summand. Since  $|\rho_0|_{BV(\Omega)}$  is a constant we retain a problem of the form:

$$\min_{y^i} \int_{\Sigma} K(\rho_0(y^i) \det(\nabla y)) - f_i \log(K(\rho_0(y^i) \det(\nabla y))) d\sigma + S^{hyper}(y^i) \quad (59)$$

If we consider the Kullback-Leibler divergence being an distance measure, this is the form of a standrad registration problem from [43] with hyperelastic regularization. Nevertheless we should mention that the Kullback-Leibler distance measure is defined on the detector domain  $\Sigma$  and not on the image domain  $\Omega$  like standard distance measures such as SSD or the Normalized Gradient Field [43]. Similar to the widely known FAIR toolbox [43] we perform a quasi Newton type optimization, which we will outline shortly in the following, starting with the discretization:

Since  $\Sigma$  is the detector domain and therefore discrete with size  $m_{\Sigma}$ , the integral becomes a sum:

$$\sum_{j=1}^{m_{\Sigma}} K(\rho_0(y^i) \det(\nabla y^i))(j) - f_i(j) \log(K(\rho_0(y^i) \det(\nabla y^i))(j)) + S^{hyper}(y(\cdot, t_i)) \quad (60)$$

Rather than classical Gauss-Newton we aim to perform a minimization with a modified BFGS method [37]. In order to minimize the objective function (60) we need to compute

the derivatives with respect to the transformation grid. We assume that the transformation is given on a nodal grid:

$$yc \in \mathbb{R}^{\tilde{m}} \quad \tilde{m} = d \prod_{i=1}^d (m_i + 1), \quad (61)$$

where  $m = (m_1, \dots, m_d)$  denotes the size of the digital image obtained in the reconstruction step. Again we use the interpolation `inter` and computation of the Jacobian determinant `jac`, implemented in the FAIR toolbox [43]. Then for any continuously differentiable distance term

$$D : (\mathbb{R}^{m_\Sigma})^2 \rightarrow \mathbb{R} \quad (K(\text{inter}(\rho^0, yc) \text{jac}(yc)), f) \rightarrow D(K(\text{inter}(\rho^0, yc) \text{jac}(yc)), f) \quad (62)$$

the derivative with respect to the transformation grid  $yc$  is given by the chain rule as

$$\begin{aligned} & \frac{d}{dyc} D(K(\text{inter}(\rho^0, yc) \text{jac}(yc)), f) \\ &= \frac{d}{dw} D(K(\text{inter}(\rho^0, yc) \text{jac}(yc)), f) \left( K \left( \frac{d}{dyc} \text{inter} + \frac{d}{dyc} \text{jac} \right) \right). \end{aligned} \quad (63)$$

With the help of (63) we can deal with different distance terms in the same objective function.

The actual registration is then performed with a multilevel approach. Since the field of view in the scanner is often bigger than the studied object, it is reasonable to impose Dirichlet boundary conditions on the motion with  $y|_{\partial\Omega} = Id$ , which guarantees the existence of a minimizer (Remark 16). This boundary conditions can be realized by taking the identity as starting guess and modifying the search directions to zero at the boundary.

## 5. RESULTS

This section is divided into two parts. In the first part we study the performance of motion and image estimation with help of a simple artificial deblurring problem. After having shown the superiority of the proposed method on this dataset we turn our focus towards a proof of concept for the clinical applicability by studying the XCAT software phantom.

**5.1. Artificial Deblurring Example.** Deblurring problems are often occurring in microscopy, where the exact image gets convoluted with an (often unknown) point-spread function [36, 14]. Since the focus in this section lies on the motion-corrected reconstruction, we assume the exact blur operator to be known. In order to assess the performance of the proposed method we consider a ring shaped object in three different stages of shrinkage (Figure 1a).

As we see in Figure 1b the proposed method performs better than the TV-regularized expectation maximization. To conclude this quantitative comparison we focus on different motion estimation methods. In order to assess the quality of the estimation we compare the motion estimation by the proposed method with an affine mass-preserving 2D registration performed on TV regularized single gate reconstructions. For doing so, we picked the best reconstruction in the terms of the reconstruction error for the EM-TV reconstruction of each gate and registered them.

We see that the inaccurate motion estimation by registering TV regularized single gate reconstruction has severe impact on the reconstruction quality. Since the edges are not exactly aligned by the registration, this imprecise motion estimation leads to an inexact size of the ring in the reconstruction, respectively to a blurring of the edges in the averaging process (Figure 3).

In conclusion we can state the advantages of the proposed method are twofold. First incorporating motion-estimation directly is superior to averaging methods, since the reconstruction is performed from full data instead of just averaging images reconstructed from parts of the data. Despite small errors in the motion information the proposed methods also performs better than the single gate reconstruction with TV regularization. Additionally the motion-estimation yielded by the proposed method was clearly superior to the registration based methods. We can give the following three possible explanations for this:

- In contrast to single gate reconstructions the proposed methods transforms a template image, which is generated from the full data set.
- By projecting the transformed template into the measurement domain the error occurring from reconstructing the reference image from part of the data is avoided.
- The Kullback-Leibler data fidelity used in the proposed method is directly adapted to the Poisson noise characteristics of the data, while the SSD distance we use for the registration is used for Gaussian denoising.

**5.2. XCAT Software Phantom.** In this subsection we inspect the performance of three reconstructions methods on data generated by the XCAT software phantom [52]. For generating the data we projected four cardiac gates of the phantom into the data spaces specified by the Siemens Biograph Sensation 16 scanner provided by the EMRecon toolbox [33]. The projected data was downsampled by the factor 1000 corrupted with Poisson noise and then scaled up again. We reconstructed the first gate from this single gate data with the classical EM algorithm, EM-TV algorithm, and the method we described in this chapter applied on the full data.

Visual inspection of the reconstructed images shows that despite having a higher reconstruction error the proposed method captures structures with a low amount of activity better than both single gate reconstruction methods (Figure 4). As we see the right ventricle stands out much clearer by incorporating motion information. This illustrates that the proposed method shows its potential for regions with a poor signal-to-noise ratio.

## 6. WEAK FORMULATION OF THE RECONSTRUCTION PROBLEM

The analysis of functional (5) showed, that we needed assumptions on the injectivity resp. boundedness of the Banach indicatrix of the transformation in order to derive analytical results. We want to outline a weak formulation, which guarantees mass preservation, but does not assume injectivity of the transformation. We conclude this outlook by showing that the weak formulation implies injectivity for the motion  $y$  and thus corresponds to restricting the admissible set of transformation to a certain subset of the weak diffeomorphisms.

We start by giving an equivalent formulation of the transformation model (2) for a diffeomorphism  $y^i$  with inverse  $z^i$ :

$$\int_{\Omega^i} \rho^i(x) \varphi(x) dx = \int_{\Omega^i} \rho^0(y^i(x)) \det(\nabla y^i(x)) \varphi(x) dx \quad \forall \varphi \in C_0(\Omega). \quad (64)$$

By applying the change of variables formula we obtain

$$\int_{\Omega^i} \rho^i(x) \varphi(x) dx = \int_{\Omega} \rho^0(x) \varphi(z^i(x)) dx \quad \forall \varphi \in C_0(\Omega). \quad (65)$$

Since  $\rho^i \geq 0$  we can relax to Radon measures with  $d\mu^i$  generalizing  $\rho^i dx$ :

$$\int_{\Omega^i} \varphi(x) d\mu^i(x) = \int_{\Omega} \rho^0(x) \varphi(z^i(x)) dx \quad \forall \varphi \in C_0(\Omega). \quad (66)$$

Alternatively we can use  $\varphi(y^i(x))$  as test function:

$$\int_{\Omega^i} \varphi(y^i(x)) d\mu^i(x) = \int_{\Omega} \rho^0(x) \varphi(x) dx \quad \forall \varphi \in C_0(\Omega). \quad (67)$$

In order to use (66) as a constraint for minimizing (5) we provide consistency with the strong formulation by the following proposition.

**Proposition 25.** *Let  $\rho^0$  be nonnegative and  $z^i \in \mathcal{A}$  an admissible transformation. Then there exists a unique, nonnegative  $\mu^i$  satisfying (66), which fulfils the mass-preservation property*

$$\int_{\Omega^i} d\mu^i(x) = \int_{\Omega} \rho^0(x) dx. \quad (68)$$

*Proof.* Note that for any fixed  $\rho^i$  and  $z^i$  the right hand side of (66) defines a linear functional. Thus there exists a unique Radon measure  $\mu^i$  satisfying (66) [25, Chapter 1.4 Theorem 2]. The mass-preservation condition (68) can be shown by using test functions converging to constants.  $\square$

As a consequence we obtain for the weak formulation of the motion-corrected reconstruction problem:

$$\begin{aligned} \tilde{J}(\mu, z) &= \sum_{i=0}^N (D(K\mu^i, f^i) + \alpha^i |\mu^i|_{BV(\Omega)}) + \sum_{i=1}^N \beta^i S^{hyper}(z^i) \\ &\text{subject to (66)} \end{aligned} \quad (69)$$

We will not elaborate further on the analysis for the weak formulation, but instead present a consistency result:

**Proposition 26.** *Let  $\rho^i \in L^1$  and  $y^i, z^i$  fulfil (64) and (66). Then we have  $N_{y^i}(\cdot, \Omega) \leq 1$  a.e. and  $z^i(x) = (y^i)^{-1}(x)$  for a.e.  $x \in \Omega$ .*

*Proof.* We can deduce with the second part of the area formula (Theorem 6):

$$\begin{aligned}
\int_{\Omega^i} \rho^i(x) \varphi(x) dx &= \int_{\Omega^i} \rho^0(y^i(x)) \det(\nabla y^i(x)) \varphi(x) dx \\
&= \int_{\mathbb{R}^d} \sum_{w \in ((y^i)^{-1}(x) \cap \Omega)} \rho^0(y^i(w)) \varphi(w) dx \\
&= \int_{\mathbb{R}^d} \rho^0(x) \sum_{w \in ((y^i)^{-1}(x) \cap \Omega)} \varphi(w) dx.
\end{aligned}$$

By using Proposition 25 and (66) we obtain:

$$\int_{\mathbb{R}^d} \rho^0(x) \sum_{w \in ((y^i)^{-1}(x) \cap \Omega)} \varphi(w) dx = \int_{\Omega} \rho^0(x) \varphi(z^i(x)) dx \quad \forall \varphi \in C_0(\Omega). \quad (70)$$

By using test functions  $\varphi$  converging to constants, we can now deduce, that  $N_{y^i}(\cdot, \Omega) \leq 1$ . It follows that  $y^i$  is weakly invertible and we can deduce

$$\int_{\mathbb{R}^d} \rho^0(x) \varphi((y^i)^{-1}(x)) dx = \int_{\Omega} \rho^0(x) \varphi(z^i(x)) dx \quad \forall \varphi \in C_0(\Omega). \quad (71)$$

Since (71) holds for all test functions  $\varphi \in C_0(\Omega)$ , the assertion follows.  $\square$

The proposition above guarantees that for a solution  $(\mu, z)$  of the weak formulation (69), which can be expressed via (66) and (64), the motion field  $z^i$  is the weak inverse of  $y^i$ . Whether such an  $y$  exists for a given weak solution  $(\mu, z)$  is object to further research as well as a detailed analysis for the weak formulation.

## 7. DISCUSSION AND OUTLOOK

We have presented a novel variational approach to motion corrected reconstruction of density images. After the motivation of the model with help of Bayesian statistics, we proceeded to the analysis of the model. Central part of the analysis was Theorem 18, which ensured us weak  $L^1$ -convergence of the sequence of transformed images, where images as well as transformations were sequences. To prove this theorem we relied heavily on regularity properties granted by the regularizers for intensity and motion vectors. A critical point was the boundedness of the Banach indicatrix in  $L^\infty$ . It remains unclear, if we can establish a bound of the form

$$\|N_y(\cdot, \Omega)\|_\infty \leq C(\Omega) \mathcal{S}^{hyper}(y). \quad (72)$$

In order to give existence results independent of this assumption, we restricted ourselves to the case of injective transformations. Although this restriction can be motivated by a mass-preservation demand, which is in our model only granted for injective transformations, the numerical realization can again be challenging. We proposed a framework with injective Dirichlet-boundary conditions to guarantee injectivity for the whole domain, but an extension to a less restrictive model might be of interest (see e.g. [3] for a diffeomorphic registration framework).

We thoroughly tested the method on an artificial deblurring example and showed superiority to several other reconstruction methods. By applying the method on software-phantom data we gave a proof-of-concept for the applicability for real data. Although there are still difficulties to deal with the movement of really small objects the method seems to be well-suited for motion-corrected reconstruction of clinical data, especially with a low signal-to-noise ratio. Again a thorough evaluation on clinical data is the next step to go, as well as finding new means to assess the quality of a found transformation in motion correction.

Motion-corrected reconstruction with alternating minimization has been done e.g. by Mair et al. [41], although neither the concept of mass-preservation nor TV regularization is imposed. To the best of our knowledge the work of Blume et al. [8] is most closely related to the presented framework: In [8] the authors propose a similar general framework to reconstruction with simultaneous motion estimation and incorporate a local invertibility constraint in [9]. While the focus of Blume et al. lies on the actual implementation of a reconstruction method with a parametric B-spline transformation model, our main contribution are:

- A thorough analysis of the motion-corrected reconstruction problem with mass-preserving transformation model and appropriate regularization for density image and motion.
- A framework for the numerical solution of the motion corrected reconstruction problem.

In the following we present some open questions, which can motivate further research in this field.

- Incorporation of attenuation correction: In general there is only an attenuation map for one gate available, so it is a possibility to deform said map for the reconstruction of the other gates. Challenges lie in the implementation as well as the analysis.
- Incorporating of a priori information into the reconstruction framework. This can be either structural information via MR images [16] or information on the motion as boundary values [42] or estimated motion from other modalities [46].
- Another open problem are the convergence properties of the proposed alternating minimization algorithm: Due to Beck [7] we can guarantee that the sequence generated by our algorithm has a stationary point as accumulation point, if we impose some reasonable conditions on discretization and interpolation. Although we cannot expect convergence to a global minimum due to the nonconvexity of the problem, proximal regularized minimization algorithms [4] might improve the performance.

## APPENDIX A. WEAK DIFFEOMORPHISMS

We will present some basic ideas and theorems from [23], how to generalize diffeomorphisms for Sobolev mappings, which are not necessarily differentiable. Giaquinta's central idea for a mapping

$$y : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^m \tag{73}$$



is to use properties of the graph  $\mathcal{G} \subset \mathbb{R}^d \times \mathbb{R}^m$ . Based on this idea he introduced Cartesian currents [21]. As the description of these currents is beyond the scope of this paper, we will only present the results from [23] related to weak diffeomorphisms and refer to [23, 25, 27] for a detailed course on Cartesian currents. Because we do not want to discuss the theory on Cartesian currents, we define a sufficient class of transformations, which contains transformations fulfilling some rather complicated requirements from the theory of Cartesian currents.

**Definition 27.** *We define the following two subclasses of Sobolev mappings:*

$$\begin{aligned} A_{p,q}(\Omega) &:= \{y \in W^{1,p}(\Omega; \mathbb{R}^m) \mid \text{cof}(\nabla y) \in L^q\}, \\ A_{p,q}^+(\Omega) &:= \{y \in A_{p,q}(\Omega) \mid \det(\nabla y) > 0 \quad \text{a.e.}\}. \end{aligned}$$

Next we introduce weak inverses as in [23]:

**Definition 28** (Weak inverse). *Given a measurable map*

$$y : \Omega \rightarrow \hat{\Omega} \quad \lambda(\Omega) > 0 \quad \lambda(\hat{\Omega}) > 0. \quad (74)$$

*We say that*

1. *y is weakly invertible with weak inverse  $\hat{y}$ , if and only if*

$$\hat{y}(y(x)) = x \quad \text{for almost every } x \in \Omega, \quad (75)$$

$$y(\hat{y}(z)) = z \quad \text{for almost every } z \in \hat{\Omega}. \quad (76)$$

2. *y is a weak one-to-one transformation, iff there exists a measurable map*

$$\hat{y} : \hat{\Omega} \rightarrow \Omega,$$

*such that*

- a) *y and  $\hat{y}$  fulfill Lusin's condition (16).*
- b) *y and  $\hat{y}$  are the inverses of the respective other.*

The next theorem provides some properties of the inverse of a mapping:

**Theorem 29.** *Let  $y : \Omega \rightarrow \hat{\Omega}$  be a weakly invertible map with inverse  $\hat{y}$ . Suppose that*

- (i) *y satisfies Lusin's condition (16),*
- (ii) *y is almost everywhere approximately differentiable in  $\Omega$ .*

*Then  $\hat{y}$  is approximately differentiable almost everywhere in  $\hat{\Omega}$ . Moreover:*

$$Dy(\hat{y}(z))D\hat{y}(z) = Id_{\hat{\Omega}} \quad \text{for a.e. } z \in \hat{\Omega} \quad (77)$$

$$D\hat{y}(y(x))Dy(x) = Id_{\Omega} \quad \text{for a.e. } x \in \Omega \quad (78)$$

*Proof.* See [23, Chapter 3, Theorem 2]. □

With this at hand Giaquinta et al. [23] define global invertibility by using properties of the graph of a map.

**Definition 30** (Global invertibility for a.e. approximately differentiable mappings). *Let y be an a.e. approximately differentiable map from  $\Omega \subset \mathbb{R}^d$  into  $\mathbb{R}^m$  with  $\det(Dy) \in L^1(\Omega)$  satisfying*

$$\det(Dy) > 0 \quad \text{a.e. in } \Omega. \quad (79)$$

We say  $y$  is globally invertible if and only if

$$\int_{\Omega} \phi(x, y(x)) \det(Dy(x)) dx \leq \int_{\mathbb{R}^m} \left( \sup_{x \in \Omega} \phi(x, z) \right) dz \quad (80)$$

holds for all  $\phi \in C_c^0(\Omega \times \mathbb{R}^m)$  with  $\phi \geq 0$ .

Consequently we denote the set of injective functions by  $\mathcal{I}(\Omega)$ .

The definition above is related to the area formula; for any function  $\phi \in C_c^0(\Omega \times \mathbb{R}^m)$  we observe:

$$\int_{\Omega} \phi(x, y(x)) \det(Dy)(x) dx \leq \int_{\Omega} \underbrace{\sup_{x \in \Omega} \phi(x, y(x)) \det(Dy)(x)}_{=: \psi(y(x))} dx = \int_{\mathbb{R}^m} \sup_{x \in \Omega} \phi(x, z) N_y(z, \Omega) dz$$

This illustrates, how the invertibility condition (80) can be violated by functions which are not injective on sets with positive measure. According to [23] global invertibility can be defined equivalently in several other ways:

**Proposition 31.** *Let  $y : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$  be a.e. approximately differentiable in  $\Omega$  with  $\det(Dy) \in L^1(\Omega)$  and  $\det(Dy) \geq 0$  a.e. in  $\Omega$ . Then the following claims are equivalent:*

- (i)  $y$  is globally invertible.
- (ii) For any  $\phi \in C_c^0(\mathbb{R}^m)$  with  $\phi \geq 0$   $y$  satisfies the inequality

$$\int_{\Omega} \phi(y(x)) \det(Dy)(x) dx \leq \int_{\mathbb{R}^m} \phi(z) dz. \quad (81)$$

- (iii) For almost every  $z \in \mathbb{R}^m$  we have

$$N(y, \Omega, z) \leq 1. \quad (82)$$

- (iv) For almost every  $z \in \mathbb{R}^m$  we have

$$N_y(\Omega, z) = \chi_{y(\Omega)}(z) := \begin{cases} 1 & z \in y(\Omega) \\ 0 & z \notin y(\Omega) \end{cases}. \quad (83)$$

- (v) We have

$$\int_{\Omega} \det(Dy)(x) dx = \mathcal{H}^m(\tilde{y}(\Omega)) \quad (84)$$

where  $\mathcal{H}^m$  is the  $m$ -dimensional Hausdorff measure and  $\tilde{y}$  a Lusin representative.

- (vi) The inequality

$$\int_{\Omega} \det(Dy)(x) dx \leq \mathcal{H}^m(y(\Omega)) \quad (85)$$

holds for any representative of  $y$ .

*Proof.* See [23, Chapter 5, Proposition 1] □

Having this in mind we can now define a norm for the class of weak diffeomorphisms:

**Definition 32** (Norm for weak diffeomorphisms). *For any almost everywhere approximately differentiable map  $y : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$  we set*

$$|M(Dy)| := (1 + |Dy|^2 + |\operatorname{cof}(Dy)|^2 + |\det(Dy)|^2)^{\frac{1}{2}} \quad (86)$$

and define

$$\|y\|_{\operatorname{dif}^{p,q}} := \int_{\Omega} \left( |y|^p + |M(Dy)|^p + \frac{|M(Dy)|^q}{|\det(Dy)|^{q-1}} \right) dx. \quad (87)$$

Now we can define the class of weak diffeomorphisms:

**Definition 33** (Space of weak diffeomorphisms). *We say that a map  $y : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$  belongs to the class  $\tilde{\operatorname{dif}}^{p,q}(\Omega, \mathbb{R}^m)$  for  $p, q \geq 1$ , if and only if:*

- (1)  $|M(Dy)| \in L^p$ .
- (2)  $y$  has a closed graph in  $\Omega \times \mathbb{R}^m$ .
- (3)  $\det(Dy) > 0$  a.e. in  $\Omega$ .
- (4)  $y$  is globally invertible.
- (5)  $\|y\|_{\operatorname{dif}^{p,q}} < \infty$ .

A map only fulfilling the first three conditions is called a weak local diffeomorphism.

**Remark 34.** *Since giving a thorough definition of the second property requires insight in the theory of Cartesian currents, we will not elaborate further on this subject. Details can be found e.g. in [21, 25, 27]. However we can state that  $y \in A_{d-1, \frac{d}{d-1}}$  is sufficient [24] but not necessary [26] to guarantee the closedness of the graph.*

## REFERENCES

- [1] H. W. ALT, *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*, Springer Lehrbuch, Springer, 2002.
- [2] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, Oxford, New York, 2000.
- [3] J. ASHBURNER, *A fast diffeomorphic image registration algorithm*, NeuroImage, 38 (2007), pp. 95–113.
- [4] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality*, Mathematics of Operations Research, 35 (2010), pp. 438–457.
- [5] H. ATTOUCH, G. BUTTAZZO, AND G. MICHAILLE, *Variational Analysis in Sobolev and BV Spaces*, MPS-SIAM Series on Optimization, 2006.
- [6] S. BANACH, *Sur les lignes rectifiables et les surfaces dont l'aire est finie*, Fundamenta Mathematicae, 7 (1925), pp. 225–236.
- [7] A. BECK, *On the Convergence of Alternating Minimization for Convex Programming with Applications to Iteratively Reweighted Least Squares and Decomposition Schemes*, SIAM Journal on Optimization, 25 (2015), pp. 185–209.
- [8] M. BLUME, A. MARTINEZ-MÖLLER, A. KEIL, N. NAVAB, AND M. RAFECAS, *Joint reconstruction of image and motion in gated Positron Emission Tomography*, IEEE Transactions on Medical Imaging, 29 (2010), pp. 1892–1906.
- [9] M. BLUME, N. NAVAB, AND M. RAFECAS, *Joint image and motion reconstruction for PET using a B-spline motion model*, Phys Med Biol, 57 (2012), pp. 8249–8270.
- [10] L. BROUWER, *Über Abbildung von Mannigfaltigkeiten*, Mathematische Annalen, 71 (1911), pp. 97–115.

- [11] C. BRUNE, A. SAWATZKY, AND M. BURGER, *Bregman-EM-TV methods with application to optical nanoscopy*, in Scale Space and Variational Methods in Computer Vision, X.-C. Tai, K. Mrken, M. Lysaker, and K.-A. Lie, eds., vol. 5567 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2009, pp. 235–246.
- [12] M. BURGER, J. MODERSITZKI, AND L. RUTHOTTO, *A hyperelastic regularization energy for image registration*, SIAM Journal on Scientific Computing, 35 (2013), pp. B132–B148.
- [13] M. BURGER AND S. OSHER, *A Guide to the TV Zoo*, in Level Set and PDE Based Reconstruction Methods in Imaging, Lecture Notes in Mathematics, Springer International Publishing, 2013, pp. 1–70.
- [14] M. CARLAVAN AND L. BLANC-FÉRAUD, *Sparse Poisson noisy image deblurring*, IEEE Transactions on Image Processing, 21 (2012), pp. 1834–1846.
- [15] M. DAWOOD, N. LANG, X. JIANG, AND K. SCHAFERS, *Lung motion correction on respiratory gated 3-D PET/CT images*, Medical Imaging, IEEE Transactions on, 25 (2006), pp. 476–485.
- [16] M. J. EHRHARDT, K. THIELEMANS, L. PIZARRO, D. ATKINSON, S. OURSELIN, B. F. HUTTON, AND S. R. ARRIDGE, *Joint reconstruction of PET-MRI by exploiting structural similarity*, Inverse Problems, 31 (2015), p. 015001.
- [17] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Mathematics and Its Applications, Springer, 1996.
- [18] L. C. EVANS, *Partial Differential Equations*, American Mathematical Society, 1998.
- [19] I. FONSECA AND W. GANGBO, *Local invertibility of Sobolev functions*, SIAM Journal on Mathematical Analysis, 26 (1995), pp. 280–304.
- [20] M. GIAQUINTA AND G. MODICA, *Mathematical Analysis: An Introduction to Functions of Several Variables*, Mathematical analysis, Birkhäuser Boston, 2009.
- [21] M. GIAQUINTA, G. MODICA, AND J. SOUCEK, *Cartesian currents, weak diffeomorphisms and existence theorems in nonlinear elasticity*, Archive for Rational Mechanics and Analysis, 106 (1989), pp. 97–159.
- [22] ———, *Area and the area formula*, Rendiconti del Seminario Matematico e Fisico di Milano, 62 (1992), pp. 53–87.
- [23] ———, *A weak approach to finite elasticity*, Calculus of Variations and Partial Differential Equations, 2 (1994), pp. 65–100.
- [24] ———, *Remarks on the degree theory*, Journal of Functional Analysis, 125 (1994), pp. 172 – 200.
- [25] ———, *Cartesian Currents in the Calculus of Variations I: Cartesian Currents*, A Series of Modern Surveys in Mathematics Series, Springer, 1998.
- [26] ———, *Some regularity properties of locally weakly invertible maps*, in Nonlinear Analysis and Continuum Mechanics, G. Buttazzo, G. Galdi, E. Lanconelli, and P. Pucci, eds., Springer New York, 1998, pp. 45–59.
- [27] ———, *Cartesian Currents in the Calculus of Variations II: Variational Integrals*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics, Springer Berlin Heidelberg, 2010.
- [28] P. HAJLASZ, *Change of variables formula under minimal assumptions*, Colloq. Math., 64 (1993), pp. 93–101.
- [29] D. HENAO AND C. MORA-CORRAL, *Invertibility and weak continuity of the determinant for the modelling of cavitation and fracture in nonlinear elasticity*, Arch. Ration. Mech. Anal., 197 (2010), pp. 619–655.
- [30] M. JACOBSON AND J. FESSLER, *Joint estimation of image and deformation parameters in motion-Corrected PET*, in Nuclear Science Symposium Conference Record, 2003 IEEE, vol. 5, 2003, pp. 3290–3294 Vol.5.
- [31] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, vol. 160, Springer, 2005.
- [32] G. KLEIN, B. REUTTER, AND R. HUESMAN, *4D Affine registration models for respiratory-Gated PET*, in Nuclear Science Symposium Conference Record, 2000 IEEE, vol. 2, 2000, pp. 15/41–15/45 vol.2.

- [33] T. KÖSTERS, K. SCHÄFERS, AND F. WÜBBELING, *EMRecon: An expectation maximization based image reconstruction framework for emission tomography data.*, In NSS/MIC Conference Record, IEEE, (2011), pp. 4365–4368.
- [34] L. V. KOVALEV, J. ONNINEN, AND K. RAJALA, *Invertibility of Sobolev mappings under minimal hypotheses*, Annales de l’Institut Henri Poincaré (C) Non Linear Analysis, 27 (2010), pp. 517 – 528.
- [35] A. S. KRONROD, *On functions of two variables*, Uspehi Matem. Nauk (N.S.), 5 (1950), pp. 24–134.
- [36] G. LANDI AND E. LOLI PICCOLOMINI, *An Improved Newton projection method for nonnegative deblurring of Poisson-corrupted images with Tikhonov regularization*, Numerical Algorithms, 60 (2012), pp. 169–188.
- [37] D.-H. LI AND M. FUKUSHIMA, *A Modified BFGS Method and its Global Convergence in Nonconvex Minimization*, Journal of Computational and Applied Mathematics, 129 (2001), pp. 15–35.
- [38] S. M. LOZINSKIĬ, *On the indicatrix of Banach*, Vestnik Leningrad. Univ., 13 (1958), pp. 70–87.
- [39] M. LUSTIG, D. DONOHO, AND J. M. PAULY, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic resonance in medicine, 58 (2007), pp. 1182–1195.
- [40] N. N. LUZIN, *The Integral and Trigonometric Series*, PhD thesis, Moscow-Leningrad, 1915.
- [41] B. A. MAIR, D. R. GILLAND, AND J. SUN, *Estimation of images and nonrigid deformations in gated emission CT*, IEEE Transactions on Medical Imaging, 25 (2006), pp. 1130–1144.
- [42] D. MANNWEILER, J. RASCH, S. SCHMID, F. WÜBBELING, AND K. P. SCHÄFERS, *PET Reconstruction of Freely-Moving Mice Comprising Registration-Based Motion Correction of Internal Structures*, in IEEE Nuclear Science Symposium & Medical Imaging Conference, 2014.
- [43] J. MODERSITZKI, *FAIR: Flexible Algorithms for Image Registration*, SIAM, 2009.
- [44] J. MÜLLER, C. BRUNE, A. SAWATZKY, T. KÖSTERS, K. P. SCHÄFERS, AND M. BURGER, *Reconstruction of short time pet scans using Bregman iterations*, in Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE, IEEE, 2011, pp. 2383–2385.
- [45] S. MÜLLER, T. QI, AND B. S. YAN, *On a new class of elastic deformations not allowing for cavitation*, Annales de l’institut Henri Poincaré Analyse non lineaire, 11 (1994), pp. 217–243.
- [46] F. QIAO, T. PAN, J. W. CLARK, AND O. R. MAWLAWI, *A motion-Incorporated reconstruction method for gated PET studies*, Phys Med Biol, 51 (2006), pp. 3769–3783.
- [47] E. RESMERITA AND R. ANDERSSSEN, *Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems*, Math. Meth. Appl. Sci., 30 (2007), pp. 1527–1544.
- [48] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear Total Variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [49] L. RUTHOTTO, *Hyperelastic Image Registration*, PhD thesis, WWU Münster, 2012.
- [50] L. RUTHOTTO, F. GIGENGACK, M. BURGER, C. WOLTERS, X. JIANG, K. SCHÄFERS, AND J. MODERSITZKI, *A simplified pipeline for motion correction in dual gated cardiac PET*, in Bildverarbeitung für die Medizin 2012, Springer, 2012.
- [51] A. SAWATZKY, C. BRUNE, T. KSTERS, F. WÜBBELING, AND M. BURGER, *EM-TV Methods for inverse problems with Poisson noise*, in Level Set and PDE Based Reconstruction Methods in Imaging, vol. 2090 of Lecture Notes in Mathematics, Springer International Publishing, 2013, pp. 71–142.
- [52] W. P. SEGARS, G. STURGEON, S. MENDONCA, J. GRIMES, AND B. M. W. TSUI, *4D XCAT phantom for multimodality imaging research*, Medical physics, 37 (2010), pp. 4902–4915.
- [53] S. SUHR, *Variational Methods for Combined Image and Motion Estimation*, PhD thesis, University of Muenster, 2015.
- [54] A. VITUSHKIN, *On higher-Dimensional variations*, master’s thesis, Moscow, 1955.

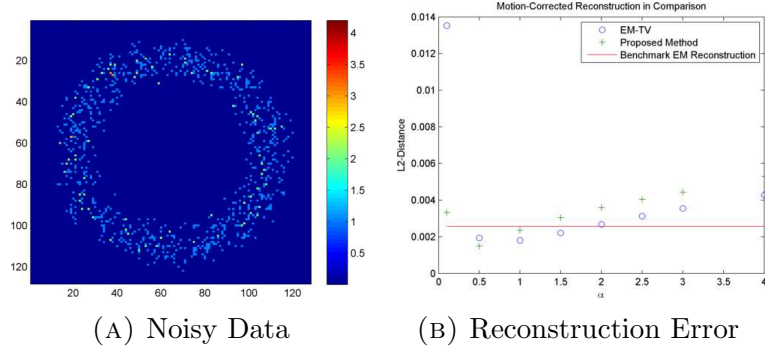


FIGURE 1. Noisy data for one Gate and reconstruction errors of the proposed Method (5 Bregman Iterations), EM-TV (5 Bregman Iterations) reconstruction and reconstruction error yielded by classical EM as benchmark. The proposed method yields the best reconstruction result.

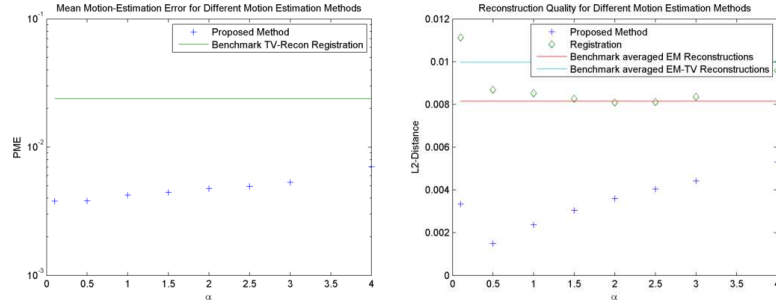


FIGURE 2. Logarithmic plot of the averaged phantom matching errors for transformations yielded by the proposed method and registration performed on the best TV regularized single gate reconstructions as benchmark. The phantom matching errors for all transformations were averaged. The error for the proposed method is smaller by an order of magnitude. Reconstruction error for the proposed method, motion-corrected EM-TV with motion determined by registration of single gate registrations and averaged (TV regularized) single gate reconstructions as benchmark. The other motion correction techniques suffer heavily from the inaccurate motion estimation.

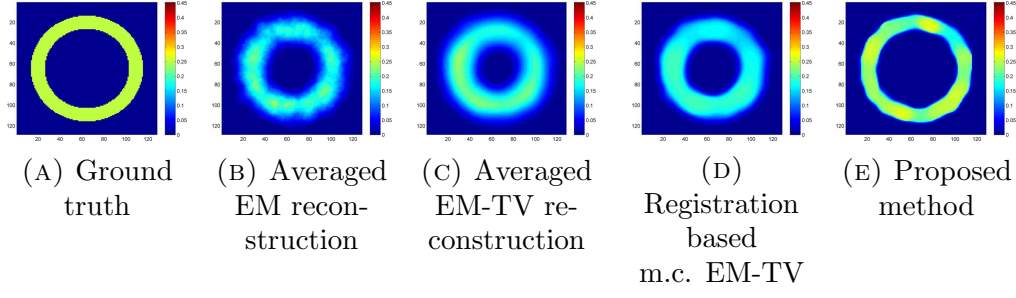


FIGURE 3. Best (TV-regularized) single gate reconstruction averaged by registration, best motion-corrected EM-TV reconstruction with motion estimation via registration of EM-TV reconstructed single gates and best reconstruction yielded by the proposed method. Note that the precise size of the ring gets lost due to the inaccurate motion estimation in the registration based methods.

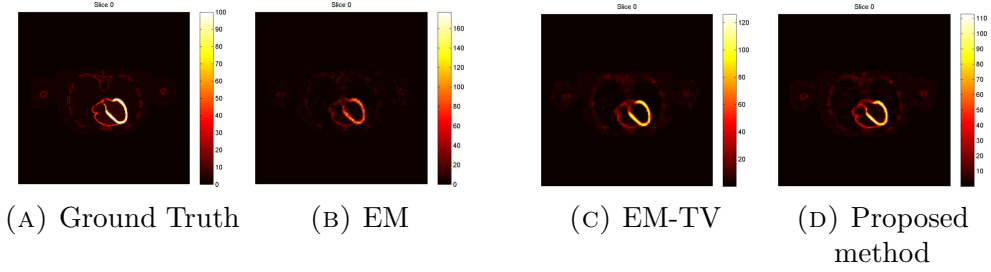


FIGURE 4. Ground truth and reconstructions EM, EM-TV ( $\alpha = 125$ ) and the proposed method ( $\alpha = 1000$ ). The TV regularized methods produce clearly better reconstruction results than the classical expectation maximization algorithm. The right ventricle is captured slightly better by the proposed method.